

A Bayesian comparison of $U(1)$ lepton flavour models

Johannes Bergström,^{1,*} Davide Meloni,^{2,†} and Luca Merlo^{3,‡}

¹ *Departament d'Estructura i Constituents de la Matèria and Institut de Ciències del Cosmos, Universitat de Barcelona, Diagonal 647, E-08028 Barcelona, Spain*

² *Dipartimento di Matematica e Fisica, Università degli Studi Roma Tre, Via della Vasca Navale 84, 00146 Roma, Italy*

³ *Instituto de Física Teórica UAM/CSIC and Departamento de Física Teórica, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain*

Whether the anarchical ansatz or more symmetric structures best describe the neutrino parameters is a long standing question that underwent a revival of interest after the discovery of a non-vanishing reactor angle and the indication of a non-maximal atmospheric angle. In this letter, a Bayesian statistical approach is adopted in order to analyse and compare the two hypotheses within the context of $U(1)$ flavour models. We study the constraints on individual model parameters and perform model comparison: the results elect constructions with built-in hierarchies among the matrix elements as preferred over the anarchical ones, with values of the evidence that depends slightly on whether the $U(1)$ charges are also considered as free parameters or not, and on the priors used.

I. INTRODUCTION

The observation of a non-vanishing reactor angle θ_{13} represents an important achievement towards a more detailed understanding of the lepton flavour sector of the Standard Model. Results on θ_{13} from reactor and accelerator experiments T2K [1–3], MINOS [4–6], DOUBLE CHOOZ [7, 8], Daya Bay [9, 10] and RENO [11] have been considered in global fits to the Pontecorvo-Maki-Nakawaga-Sakata (PMNS) matrix [12–18]. Interestingly, there is at the moment a controversial indication for deviations from the maximal value of the atmospheric angle θ_{23} , that is advocated in Refs. [16, 18], but not confirmed in Refs. [17].

The discovery of the non-vanishing reactor angle and the indication of a non-maximal atmospheric angle have a deep impact on flavour model building. Indeed, models based on discrete symmetries, that dominated the flavour scenario in the past years for their ability to describe in first approximation specific mixing patterns with $\theta_{13} = 0^\circ$ and $\theta_{23} = 45^\circ$ [19–24], need now some adjustment. A few strategies have been followed: introduction of additional parameters in preexisting minimal models; implementation of features that allow next order corrections only in specific directions in the flavour space; search for alternative mixing patterns or flavour symmetries that lead already in first approximation to $\theta_{13} \neq 0^\circ$ and $\theta_{23} \neq 45^\circ$ (see for example the reviews in Refs. [25–31] and references therein). In other words, the latest neutrino data can indeed be described in the context of discrete symmetries, but at the prize of fine-tunings and/or eccentric mechanisms.

This suggests to investigate approaches alternative to discrete flavour symmetries: from models based on con-

tinuous symmetries such as $SU(3)$ (i.e. Refs. [32–40]) or the simplest $U(1)$ (e.g., Refs. [41–45]), to models where no symmetry at all is acting on the neutrino sector (e.g., Refs. [46–49]). The latter go under the name of *anarchical* models, which have been characterised as models in which the coupling constants and mass matrices are “random” numbers drawn from an invariant probability distribution. It has been claimed that such matrices generically prefer large mixings [46–48] and that the observed sizable deviation from a zero reactor angle seems to favour anarchical models when compared to other more symmetric constructions [49]. However, as discussed in Ref. [50] for the case of neutrino mass matrices, how much a large value of a parameter is preferred can depend strongly on the definition of “preferred” and of “large”.

It has been suggested in Ref. [44] that the performances of anarchical models in reproducing the 2012 neutrino data are worse than those of models constructed upon the $U(1)$ flavour symmetry. The analysis in Ref. [44] is based on the fact that anarchy can be formulated in a $U(1)$ context, giving no charges to the left-handed fields, but non-zero $U(1)$ charges to the right-handed ones in order to describe the charged lepton mass hierarchy. This indeed allows a consistent comparison between anarchical and hierarchical models. The constructions with built-in hierarchies among the matrix elements considered in Ref. [44], that resulted to be favoured with respect to anarchical models, have been chosen due to particular phenomenological features that lead to a good description of the data. On the other hand, these models have not been shown to be necessarily the best ones available and other $U(1)$ models could provide an even better description of the data. The method of analysis used in Ref. [44] treated all mass matrix elements as random complex numbers with modulus of order one and the models were judged according to the *success probability*, i.e., the fraction of the generated points which satisfied some experimentally motivated cuts. This kind of analysis allows to naively estimate the relative success of one model with respect

*Electronic address: bergstrom@ecm.ub.edu

†Electronic address: meloni@fis.uniroma3.it

‡Electronic address: luca.merlo@uam.es

to the others, but it is not based on, nor motivated by, any established method of statistical analysis, and hence the results cannot be given any proper statistical interpretation. However, some aspects of that method are rather good approximations of some procedures having well-defined meanings within Bayesian inference. In comparison, a standard χ^2 -analysis of these models is not possible since they can all fit the data equally well for any values of the interesting parameters. The main objective of this work is to make a statistically principled analysis of the models considered in Ref. [44], to make a systematic search for even better models, and to make the appropriate generalisations, allowing for a meaningful comparison between anarchy and hierarchy in the form of $U(1)$ models. Furthermore, data on the masses of the charged leptons will be included in the analysis. This is important since this data gives constraints on the model parameters – constraints that must be consistent with neutrino oscillation data – which will result in a large impact on the final results.

In the following, basic concepts of Bayesian inference are briefly summarized in Sec. II. Section III is devoted to the presentation and analysis of specific flavour models, where the lepton $U(1)$ charges are chosen a priori due to particular features of the Yukawa matrices. In Sec. IV the lepton charges are instead treated as free parameters in the Bayesian analysis. We discuss future prospects to further distinguish among the considered models in Sec. V and present the conclusions in Sec. VI.

II. BAYESIAN INFERENCE

In the Bayesian interpretation, probability is associated with the *plausibility* or *credibility* of a proposition. Perhaps the main goal of science is to infer which model or hypothesis best describes a certain set of collected data. Also, these models should preferably be “simple” or “economical” in some sense. If one accepts the Bayesian interpretation of probability, a very powerful arsenal of inference tools become available. In a nutshell, the idea is to use the laws of probability to calculate the probabilities of different hypotheses or models, when conditioned on some known (or presumed) information.

If the collected data is denoted by \mathbf{D} and the set of considered hypotheses or models is M_1, M_2, \dots, M_r , Bayes’ theorem gives the plausibilities of each of the hypotheses after considering the data, the *posterior probabilities*,

$$\Pr(M_i|\mathbf{D}) = \frac{\Pr(\mathbf{D}|M_i) \Pr(M_i)}{\Pr(\mathbf{D})}. \quad (1)$$

Here, the *evidence* $\Pr(\mathbf{D}|M_i)$ is the probability of the data, assuming the model M_i to be true, while $\Pr(M_i)$ is the *prior probability* of M_i , which is how plausible M_i is before considering the data. $\Pr(\mathbf{D})$ is the probability of

the data without assuming any particular model.¹ One can then perform *model comparison* by calculating ratios of posterior probabilities, the *posterior odds*, of two models as

$$\frac{\Pr(M_i|\mathbf{D})}{\Pr(M_j|\mathbf{D})} = \frac{\Pr(\mathbf{D}|M_i) \Pr(M_i)}{\Pr(\mathbf{D}|M_j) \Pr(M_j)}. \quad (2)$$

In words, the posterior odds is given by the *prior odds* $\Pr(M_i)/\Pr(M_j)$ multiplied by the *Bayes factor* $B_{ij} = \Pr(\mathbf{D}|M_i)/\Pr(\mathbf{D}|M_j)$, which quantifies how much better M_i describes that data than M_j . The prior odds quantifies how much more plausible one model is than the other a priori, i.e., without considering the data. If there is no reason to favor one of the models over the other, the prior odds should be taken to equal unity (in which case the posterior odds equals the Bayes factor), but sometimes one must consider this point more carefully.

If the model contains free parameters Θ , the evidence is given by

$$\begin{aligned} \mathcal{Z} = \Pr(\mathbf{D}|M) &= \int \Pr(\mathbf{D}, \Theta|M) d^N \Theta \\ &= \int \Pr(\mathbf{D}|\Theta, M) \Pr(\Theta|M) d^N \Theta \\ &= \int \mathcal{L}(\Theta) \pi(\Theta) d^N \Theta. \end{aligned} \quad (3)$$

Here, the *likelihood function* $\mathcal{L}(\Theta) \equiv \Pr(\mathbf{D}|\Theta, M)$ is the probability (density) of the data \mathbf{D} , assuming parameter values Θ and $\pi(\Theta) \equiv \Pr(\Theta|M)$ is the prior probability (density), which should reflect how plausible different values of the parameters are, assuming the model to be correct. It should always be normalized, i.e., integrate to unity. The assignment of priors are probably the most discussed and controversial part of Bayesian inference. This assignment is often far from trivial, but constitutes a very important part of any Bayesian analysis.

One observes that the evidence is the average of the likelihood over the prior, and hence this method automatically implements a form of *Occam’s razor*, since in general a more predictive model with a smaller parameter space will have a larger evidence than a less predictive one, unless the latter can fit the data substantially better.

The probabilities of the different hypotheses give the complete posterior inference on the space of models, and these have a somewhat unique and meaningful interpretation on their own. However, Bayes factors, or rather posterior odds, are usually interpreted or “translated” into ordinary language using the so-called *Jeffreys scale*, given in Tab. I (“log” denotes the natural logarithm).

This scale has been used in applications in cosmology and astro- and particle physics such as Refs. [54–57] (and Refs. [58, 59] in neutrino physics) although slightly more

¹ However, we note that all probabilities are always conditioned on some “background information” [51–53].

$ \log(\text{odds}) $	odds	$\Pr(M_1 \mathbf{D})$	Interpretation
< 1.0	$\lesssim 3 : 1$	$\lesssim 0.75$	Inconclusive
1.0	$\simeq 3 : 1$	$\simeq 0.75$	Weak evidence
2.5	$\simeq 12 : 1$	$\simeq 0.92$	Moderate evidence
5.0	$\simeq 150 : 1$	$\simeq 0.993$	Strong evidence

TABLE I: The Jeffreys scale, often used for the interpretation of Bayes factors, odds, and model probabilities. The posterior model probabilities for the preferred model are calculated by assuming only two competing hypotheses and equal prior probabilities.

aggressive scales have been used previously [60, 61]. Note that it is often the case that the evidence is quite dependent on the prior used, although the Bayes factor will generally favour the correct model once “enough” data have been obtained.

The complete inference of the parameters within a single model is given by the posterior distribution,

$$\Pr(\boldsymbol{\Theta}|\mathbf{D}, M) = \frac{\Pr(\mathbf{D}|\boldsymbol{\Theta}, M) \Pr(\boldsymbol{\Theta}|M)}{\Pr(\mathbf{D}|M)} = \frac{\mathcal{L}(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})}{\mathcal{Z}}. \quad (4)$$

Since the evidence does not depend on the values of the parameters $\boldsymbol{\Theta}$, it is usually ignored in when estimating parameters. However, often the most interesting question does not concern the parameter values within a pre-chosen model, but rather which are the preferred ones by the data out of a given set of models.

The main result of Bayesian parameter inference is the posterior and its marginalized versions (usually in one or two dimensions). However, it is also common to give point estimates such as the posterior mean or median, as well as *credible intervals (regions)*, which are defined as intervals (regions) containing a certain amount of posterior probability. Note that these regions are not unique without further restrictions, just as for classical confidence intervals, and that in general they do not contain all the information that the posterior contains.

Although the reasoning and techniques used when performing model selection are often different than when estimating parameters, one can equally well consider model selection as a parameter inference problem with an additional discrete parameter denoting the model index. Hence, there is no real “fundamental” difference between model selection and parameter estimation. We use MULTINEST [62–64] for the evaluation of all evidences and posterior distributions in this work.

III. ANALYSIS OF SPECIFIC $U(1)$ MODELS

In this section, we review the general strategy to build $U(1)$ models, and recall the specific models previously defined and discussed in Ref. [44]. We will consider them

within the context of supersymmetry, as the holomorphicity of the superpotential simplifies the construction of the Yukawa interactions. We then show the results of the Bayesian parameter estimation and model comparison.

A. General features

The formulation of a model based on the $U(1)$ symmetry [65] in the supersymmetric context is simple and elegant:

- The flavour symmetry acts horizontally on leptons and the charges can be written as $e^c \sim (n_1^R, n_2^R, 0)$ for the $SU(2)_L$ lepton singlets and as $\ell \sim (n_1^L, n_2^L, 0)$ for the $SU(2)_L$ lepton doublets. The third lepton charges can be set to zero as only charge differences have an impact on mass hierarchies and on mixing angles. Furthermore, it is not restrictive to assume $n_1^R > n_2^R > 0$ in order to guarantee the correct ordering of the charged leptons. The Higgs fields $H_{u,d}$ are not charged under $U(1)$ to prevent flavour-violating Higgs couplings.
- Once leptons have $U(1)$ charges, the Yukawa terms are no longer invariant under the action of the flavour symmetry. To formally recover the invariance, a new scalar field (or more than one in non-minimal models) can be introduced, the flavon θ , that transforms non-trivially only under $U(1)$, with charge n_θ . Then, the Yukawa Lagrangian can be written as

$$\mathcal{L}_Y = (y_e)_{ij} \ell_i H_d e_j^c \left(\frac{\theta}{\Lambda} \right)^{p_e} + (y_\nu)_{ij} \frac{\ell_i \ell_j H_u H_u}{\Lambda_L} \left(\frac{\theta}{\Lambda} \right)^{p_\nu} + \text{H.c.} \quad (5)$$

where Λ is the cut-off of the effective flavour theory and Λ_L the scale of the lepton number violation, in principle distinct from Λ . $(y_e)_{ij}$ and $(y_\nu)_{ij}$ are free parameters: for naturalness, these parameters are taken to be complex and with modulus of order 1. p_e and p_ν are suitable powers of the dimensionless ratio θ/Λ necessary to compensate the $U(1)$ charges for each Yukawa term and therefore recover the invariance under the flavour symmetry. Without loss of generality, we can fix $n_\theta = -1$; consequently, $n_1, n_2 > 0$ to assure that the Lagrangian expansion makes sense. Here and in the following, neutrino masses are described by the effective Weinberg operator, while the extension to ultra-violate completions, such as See-Saw mechanisms, is straightforward.

- Once the flavon and the Higgs fields develop non-vanishing vacuum expectation values (VEVs), the flavour and electroweak symmetries are broken and

mass matrices arise from the Yukawa Lagrangian. In particular, the ratio of the flavon VEV $\langle \theta \rangle$ and the cut-off Λ of the effective theory defines the expanding parameter of the theory,

$$\epsilon \equiv \frac{\langle \theta \rangle}{\Lambda} < 1. \quad (6)$$

A useful parametrisation for the Yukawa matrices then follows as

$$Y_e = F_e y_e F_\ell, \quad Y_\nu = F_\ell y_\nu F_\ell, \quad (7)$$

where $F_f = \text{diag}(\epsilon^{n_{f1}}, \epsilon^{n_{f2}}, \epsilon^{n_{f3}})$. Throughout this work, and following Ref. [44], the charges will be taken to be integers, since non-integer charges can always be redefined to integers as long as it is accompanied by a suitable redefinition of the parameter ϵ .

B. Specific $U(1)$ models

The lepton charges of those models introduced in Ref. [44] where neutrino masses are described by the Weinberg operators, are given in the upper part of Tab. II. In the lower part, there are two new models that have been identified as “good” in the analysis of Sec. IV.

Model	e_R	ℓ_L
Anarchy (A)	(3,2,0)	(0,0,0)
$\mu\tau$ -Anarchy ($A_{\mu\tau}$)	(3,2,0)	(1,0,0)
Hierarchy (H)	(5,3,0)	(2,1,0)
New Anarchy (A')	(3,1,0)	(0,0,0)
New Hierarchy (H')	(8,3,0)	(2,1,0)

TABLE II: *Upper part: the models introduced in Ref. [44] and their flavour charges under $U(1)$. Lower part: models identified in the more general analysis of Sec. IV. The flavon charge is -1 , while the Higgs charge is zero.*

From the lepton charges in Tab. II, the textures for the charged leptons Y_e and neutrino Y_ν Yukawa matrices are as follows:

$$\begin{aligned} A: \quad Y_e &= \begin{pmatrix} \epsilon^3 & \epsilon^2 & 1 \\ \epsilon^3 & \epsilon^2 & 1 \\ \epsilon^3 & \epsilon^2 & 1 \end{pmatrix}, \quad Y_\nu = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \\ A_{\mu\tau}: \quad Y_e &= \begin{pmatrix} \epsilon^4 & \epsilon^3 & \epsilon \\ \epsilon^3 & \epsilon^2 & 1 \\ \epsilon^3 & \epsilon^2 & 1 \end{pmatrix}, \quad Y_\nu = \begin{pmatrix} \epsilon^2 & \epsilon & \epsilon \\ \epsilon & 1 & 1 \\ \epsilon & 1 & 1 \end{pmatrix}, \\ H: \quad Y_e &= \begin{pmatrix} \epsilon^7 & \epsilon^5 & \epsilon^2 \\ \epsilon^6 & \epsilon^4 & \epsilon \\ \epsilon^5 & \epsilon^3 & 1 \end{pmatrix}, \quad Y_\nu = \begin{pmatrix} \epsilon^4 & \epsilon^3 & \epsilon^2 \\ \epsilon^3 & \epsilon^2 & \epsilon \\ \epsilon^2 & \epsilon & 1 \end{pmatrix}, \end{aligned} \quad (8)$$

$$\begin{aligned} A': \quad Y_e &= \begin{pmatrix} \epsilon^3 & \epsilon & 1 \\ \epsilon^3 & \epsilon & 1 \\ \epsilon^3 & \epsilon & 1 \end{pmatrix}, \quad Y_\nu = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \\ H': \quad Y_e &= \begin{pmatrix} \epsilon^{10} & \epsilon^6 & \epsilon^2 \\ \epsilon^9 & \epsilon^5 & \epsilon \\ \epsilon^8 & \epsilon^4 & 1 \end{pmatrix}, \quad Y_\nu = \begin{pmatrix} \epsilon^4 & \epsilon^3 & \epsilon^2 \\ \epsilon^3 & \epsilon^2 & \epsilon \\ \epsilon^2 & \epsilon & 1 \end{pmatrix}. \end{aligned} \quad (9)$$

In the spirit of $U(1)$ models, the coefficients in front of ϵ^n are expected to be complex numbers with absolute values of $\mathcal{O}(1)$ and arbitrary phases. Considering that Y_ν is a symmetric matrix, the total number of parameters that should be considered in the analysis is 30, from the Yukawa matrices, plus the unknown value of ϵ .

In Ref. [44], the performances of the first three models, A , $A_{\mu\tau}$ and H , were evaluated by considering the fraction of the corresponding parameter spaces which were consistent (at a fixed confidence level) with the experimental constraints, for fixed values of ϵ . The main result of that analysis was that, using a uniform distribution for the $\mathcal{O}(1)$ coefficients in the interval $[0.5, 2]$ or $[0.8, 1.2]$ (and phases with a uniform distribution in $[0, 2\pi]$), H was the best performing model for values of ϵ larger than about 0.3, while for smaller values $A_{\mu\tau}$ had the best success rate. At the same time, A seemed disfavoured when compared to the previous models, for almost all the range of values of ϵ . However, it is worth to mention that i) the doublet charges of H and $A_{\mu\tau}$ were chosen in order to naturally reproduce the neutrino data; ii) these models had an intrinsic advantage over A , as they have an extra parameter in the form of ϵ , while A is insensitive to its value; iii) although the charges and the relevant values of ϵ considered were chosen using the observed charged lepton masses (as well as quark masses and mixings as these models were formulated in a $SU(5)$ GUT context), they were not subsequently used in the numerical analysis.

The charges of the remaining models, A' and H' , are chosen following the analysis in Sec. IV, where they are treated in general as free parameters. In particular, the charges of A' are identified as the best (i.e. having large posterior probabilities) in the case of vanishing doublet charges, whereas those of H' are the best assuming all free charges. They are rather similar to those of A and H , respectively.

In the remaining part of this section, we will consider the models listed in Tab. II with the aim of:

- Analyse and compare the models using Bayesian inference;
- Check whether the results of Ref. [44] remain valid when performing the full Bayesian analysis (assuming that the differences in the data sets used for our analysis and that in Ref. [44] are irrelevant for the model comparison);
- Determine the importance of including the charged lepton data.

C. Bayesian analysis and priors

In order to calculate the evidence and obtain the posterior distributions, we need to specify priors on the 31 free parameters of these models.

- It is reasonable to take ϵ as a priori independent of all the $\mathcal{O}(1)$ coefficients, phases, and charges, and then we use consistency of the lagrangian expansion to set an upper bound of 0.6. We use a prior

$$\pi(\epsilon) = \frac{N^{-1}(\epsilon_0)}{1 + \epsilon/\epsilon_0}, \quad \epsilon \in [0, 0.6], \quad (10)$$

which behaves uniformly in $\log \epsilon$ for $\epsilon \gg \epsilon_0$ and uniformly in ϵ for $\epsilon \ll \epsilon_0$. $N(\epsilon_0)$ is the required normalisation factor. We take $\epsilon_0 = 10^{-2}$ as our default choice, but we will find that our results are quite insensitive to changes in ϵ_0 , which is to be expected since it is a free parameter in all models we consider.

- For the $\mathcal{O}(1)$ parameters, we also make the reasonable assumption that they are a priori independent of the charges, and so their priors should be the same for the cases of vanishing and non-zero charges. This translates into a *a priori* invariance under basis rotations and leads to a unique measure on the leptonic mixing matrix, the so-called Haar measure. This has been studied in some detail in Refs. [46–49], and its interpretation has often been that it describes how matrix elements or mixing angles are “randomly distributed” in some sense. However, these distributions are more naturally considered in a Bayesian context, and the use of Haar measures to construct prior distributions has been extensively studied in the statistics literature (see Ref. [66] and references therein).

In addition to the mixing angles, there still remains an arbitrary measure over the neutrino mass eigenvalues. As pointed out in Ref. [67], under the additional assumption that the matrix elements are independent of each other, the measure on the mass matrix becomes unique (up to a scale),

$$\pi(m_{ij}) \propto e^{-\text{Tr}(mm^\dagger)/2}, \quad (11)$$

and so the real and imaginary parts of each element are a priori independent with Gaussian priors. Note that, since the mass matrix is symmetric, the off-diagonal elements are on average a factor $\sqrt{2}$ smaller than the diagonal elements, but this will have a negligible impact on the results. Equivalently, the prior on the absolute value q and phase ϕ of each (off-diagonal) element is

$$\pi(q, \phi) = \frac{qe^{-q^2/2}}{2\pi}, \quad (12)$$

with mode at $q = 1$.

By analogy, we take the same prior for the elements of the charged lepton mass matrix, although, since it is not symmetric, the elements have the same widths of their priors.

D. Data

We consider the following relevant data:

- Neutrino oscillation data constrain the parameters $r = \Delta m_{21}^2/\Delta m_{31}^2$, s_{12}^2 , s_{23}^2 , and s_{13}^2 (using s_{ij} for $\sin \theta_{ij}$). Although there are some constraints on the CP-violating phase δ , we will not take this into account as induced priors on δ are rather similar and independent of ϵ in all models, and this will have negligible impact on the results. The oscillation parameters are rather well constrained and the correlations between the oscillation in the standard parameterization are rather small, and so we can approximate the oscillation likelihood as

$$\mathcal{L}_{\text{osc}}(\Theta) \simeq \mathcal{L}^1(r)\mathcal{L}^2(s_{12}^2)\mathcal{L}^3(s_{23}^2)\mathcal{L}^4(s_{13}^2). \quad (13)$$

We take the individual likelihood components as Gaussian functions using the results of Ref. [17]. Although this might not be a perfect approximation, this should not have a noticeable impact on any results.

- The ratio of the charged lepton masses, $m_\tau/m_e \simeq 3477$ and $m_\mu/m_e \simeq 207$ have been measured with very good accuracy, and so we can approximate the associated likelihoods with Dirac δ -functions. To allow convergence of our numerical analysis we will in turn approximate these likelihoods using rather broad Gaussians. This will not change any conclusions as long as the widths of these Gaussians are taken small enough (so that all priors are effectively constant over these widths). We take our errors to be about 3–5%, and have checked numerically that our results are insensitive to changes in these widths.
- In what follows, only the normal mass ordering for the neutrino spectrum will be considered, as this is much strongly preferred than the inverted one. This is a well known result and already pointed out in Ref. [42]: the inverted mass ordering is typically linked to a maximal value of the solar angle, in contrast with the observations.

E. Results: constraints on ϵ

We first focus on the posterior distributions in all the models. However, since ϵ is really the only parameter of interest, and its marginal posteriors are at least not very far from being Gaussian, we only consider the posterior

means and standard deviations (which closely matches the posterior medians and 68% credible intervals).

The inclusion of the charged lepton data is particularly important since, by themselves, they can give strong constraints on ϵ . Consequently, the same preferred values of ϵ must be used when fitting the neutrino data. In general, we expect naive estimates as

$$\epsilon \simeq \left(\frac{(y_e)_{33} m_e}{(y_e)_{11} m_\tau} \right)^{1/(n_1^L + n_1^R)} \simeq \left(\frac{(y_e)_{33} m_\mu}{(y_e)_{22} m_\tau} \right)^{1/(n_2^L + n_2^R)}, \quad (14)$$

where $(y_e)_{ij}$ are the $\mathcal{O}(1)$ factors entering the Yukawa matrix as defined in Eq. (5). If each of these $\mathcal{O}(1)$ factors is identified with the absolute value of the matrix element in Eq. (12), implying a prior uncertainty of $\sigma_{\log q} \simeq 0.64$, then one can obtain naive analytical estimates of $\log \epsilon$ and its uncertainty, which can be compared with the numerical results. Note that each of the two $\mathcal{O}(1)$ factors gives a contribution to the uncertainty of ϵ . Figure 1 shows these naive estimates (in green-circles and black-diamonds, respectively) together with the numerical results when using only neutrino oscillation data (in red-crosses) and finally when also charged lepton data are considered (in blue-squares). The error bars are given by twice the naive estimates and the posterior standard deviations, respectively. In the anarchical models there are, of course, no constraints on ϵ from neutrino data.

The two naive estimates are relatively consistent with each other and with the ones using neutrino data in all the models. The largest tension can be found in the model A between the electron and muon estimates, but it is still naively smaller than 3σ .² Furthermore, the constraints using all the data is generally consistent with being the naive combination of the three other sets of data (i.e. when treated as independent measurements), for both the best estimates and the size of the errors. For some of the models the combined uncertainty is somewhat smaller than the naive expectation, and for the model H the posterior mean is smaller than all the three partial estimates. However, the naive combination is expected to be valid only when all the individual constraints are Gaussian and there are no common nuisance parameters. In the present case, we do see some non-Gaussian features of the posteriors, and there are 30 common nuisance parameters whose correlations can invalidate the naive combination.

Finally, we note that no constraints on ϵ can be obtained in a standard χ^2 -analysis in which the χ^2 is minimized (or likelihood maximized) over the remaining parameters. All the models can fit the data equally well, i.e., perfectly, for any non-zero values of ϵ . However, depending on the value of ϵ , the fit would require more or

less fine-tuning among the $\mathcal{O}(1)$ parameters. This fine-tuning is automatically considered in the Bayesian analysis and is what yields the above constraints on ϵ .

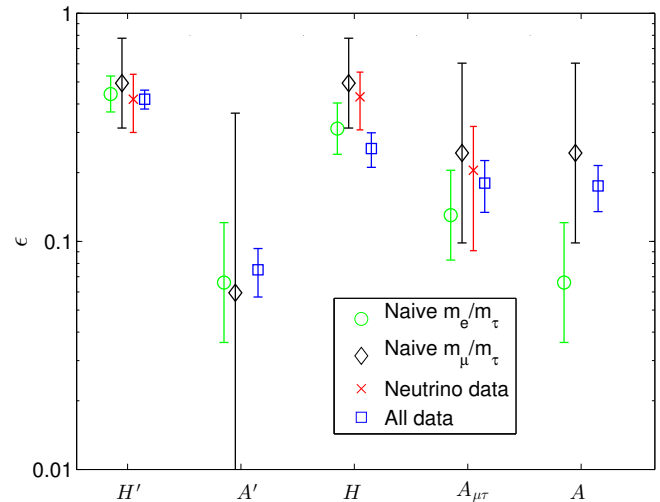


FIG. 1: Constraints on ϵ in the different models, with the error bars being twice the standard uncertainties. See text for the description of the naive estimates. For the numerical estimates (red and blue), the points are the posterior means and the uncertainties are the posterior standard deviations.

F. Results: model comparison

All the models have a χ^2 -minimum of zero and therefore a χ^2 -analysis can never “exclude” any of the models, and moreover two models cannot be meaningfully compared. The different models require more or less fine-tuning among the $\mathcal{O}(1)$ parameters, and here also of ϵ , and this aspect is automatically considered in the Bayesian analysis discussed here below.

We note that if M_c denotes the model with fixed charges $c = (n_1^R, n_2^R, n_1^L, n_2^L)$, then

$$\frac{\Pr(M_{c_1}|\mathbf{D})}{\Pr(M_{c_2}|\mathbf{D})} = \frac{\mathcal{Z}_{c_1} \Pr(M_{c_1})}{\mathcal{Z}_{c_2} \Pr(M_{c_2})}. \quad (15)$$

If there is no reason that a particular set of charges is a priori more plausible than any other, one sets $\Pr(M_{c_1})/\Pr(M_{c_2}) = 1$ and uses the Jeffreys scale to interpret the strength of evidence.

The logarithms of the evidences of the models in Tab. II normalised to the evidence of A' , i.e., the Bayes factor between all the models and A' , are reported in Fig. 2, and this allows a quantitative comparison of all those models. The numerical (statistical) uncertainty on the individual log-evidence estimates, as reported by MULTINEST, is about 0.15 for all the models. Hence, the uncertainty on the logarithms of the Bayes factors are about 0.2.

We first focus on the models A , $A_{\mu\tau}$ and H that have been previously discussed in Ref. [44]. When using only

² When estimating ϵ within this model, these discrepant data can still be consistently combined, but this tension will work against the model A when it is compared to other models.

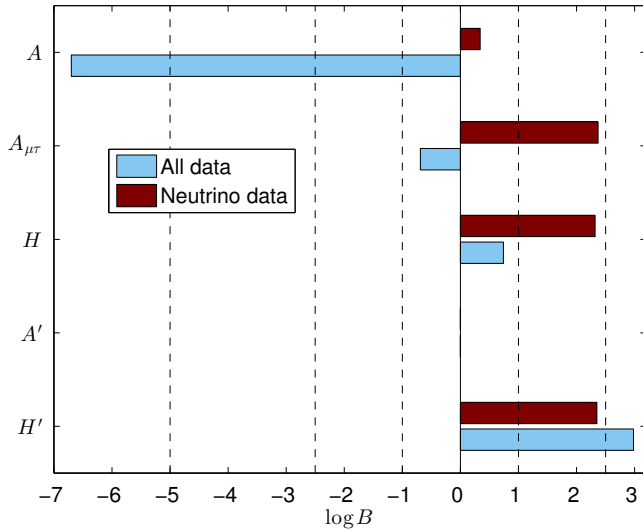


FIG. 2: Logarithms of Bayes factors with respect to the model A' for the models in Tab. II using only neutrino data (dark-red bars) and all data (light-blue bars). The dashed lines correspond to the boundaries given in Tab. I in the comparison with A' .

the neutrino data, we see from Fig. 2 that the hierarchical models are all weakly preferred compared to the anarchical ones. These results are in line with the ones quoted in Ref. [44], in the sense that the anarchical model is less appropriate to describe the data. However, this depends somewhat on the specific choice for the prior on ϵ : while the evidence of A are independent on the value of ϵ_0 , those of the hierarchical models are not. Taking, say, ϵ_0 as small as 10^{-4} would reduce the $\log B$'s of $A_{\mu\tau}$ and H by 0.7 units. This can be considered the “punishment” that the hierarchical models receive for their advantage of having an additional parameter, as already discussed in Sec. III B.

When also the charged lepton data is included, we notice that A is strongly disfavoured compared to the other models since it predicts the charged lepton data rather badly.

Considering in the comparison also the H' and A' models, H' is the best model: it is moderately better than $A_{\mu\tau}$ and A' , and weakly preferred over H . The different evidences of A and A' when only neutrino data are taken into account, are due to statistical fluctuations of the evidence estimates, but the difference is consistent with being zero within the uncertainties.

The charged lepton data has again a deep impact on the analysis: once considered, ϵ is well-constrained in all models, which implies that there is basically no sensitivity to the volume of the prior any more, although a small dependence on the shape of the prior could still remain. For example, when the prior becomes effectively uniform (taking ϵ_0 defined in Eq. (10) very large), models which subsequently prefer large values of ϵ are favoured by roughly the log of the ratio of the preferred values of

ϵ 's. In this case, the evidences for H' and H get stronger than the values shown in Fig. 2:

$$\begin{aligned} H' &\rightarrow \log B \simeq 4.5, \\ H &\rightarrow \log B \simeq 2.5. \end{aligned} \quad (16)$$

IV. GENERALISED $U(1)$ MODELS

The models A , $A_{\mu\tau}$ and H discussed in the previous section were taken as in Ref. [44] were the charges have been motivated by the data. However, one could argue that in some sense the data are used twice - first to choose the charges and then to analyse the models. On the other hand, the charges can be thought of as just a set of additional unknown free parameters, which should preferably not be fixed from the beginning, but instead inferred in the Bayesian analysis. In this section, this latter strategy will be followed to study parameter constraints and make a more general comparison of anarchy vs. hierarchy.

Beside the 31 free parameters considered when the charges are fixed, the most general model with free charges has four additional discrete parameters, which we can assign priors and subsequently calculate the posteriors of in the usual way. We consider the following generalised models:

\mathcal{A}_G : Generalised anarchy. The doublets have no charges and only the $SU(2)_L$ singlet charges are non-vanishing and are taken as free parameters.

\mathcal{H}_G : Generalised hierarchy. This is the most general model where all the charges are free. Potentially, this model can give the best predictions in both the charged lepton and neutrino sectors. On the other hand, it has a larger parameter space and hence more regions in which it could fail to predict the experimental data. Hence, this model could as a whole have small predictive power. We include only non-vanishing doublet charges, so that \mathcal{A}_G is not contained within setup. This will have negligible impact on our results (except of this specific region in parameter space).

In other words, \mathcal{A}_G is the union of all models with anarchy in the neutrino sector, while \mathcal{H}_G is the union of all models with hierarchy.

A. Priors on the charges

In order to perform a Bayesian analysis with free charges, one needs to assign them priors. The doublet charges should be independent on the singlet charges, but the charges of the same field must be dependent, since it holds that $n_1 \geq n_2 > 0$. A general prior can be written as

$$\pi(n_1, n_2) = \pi(n_1|n_2)\pi(n_2). \quad (17)$$

The charges are integers, and a naturalness criterion can be introduced such that the preference falls on the set with smaller charges: very large charges should not be equally plausible as small ones, but at the same time it is wise to not assign a prior of exactly zero to a particular charge. It seems reasonable to expect $\langle n_2 \rangle = \lambda$, with $\lambda \simeq 1 - 3$. Then the unique distribution on the non-negative reals which has maximum entropy (“minimum information”, see, e.g., Refs. [51, 52]) satisfying this constraint is the geometric distribution, $\pi(n_2) = (1 - p)^{n_2} p$, with $p = 1/(\lambda + 1)$. If the same condition is imposed on $n_1 - n_2$, n_1 is expected to be λ larger than n_2 and then the prior on both charges simply becomes

$$\pi(n_1, n_2) = (1 - p)^{n_1} p^2. \quad (18)$$

Notice that this prior only depends on the value of the largest charge and the possible dependence of the results on the value of λ will be discussed in the next sections.

On the other side, a uniform prior $\pi(n_1, n_2) = \text{const.}$ for charges with $n_{\max} \geq n_1 \geq n_2$ will also be considered.

B. Results: parameter constraints

It is possible now to analyse the models with free charges and compare them using the Bayesian evidence. The posteriors of the parameters of \mathcal{A}_G and \mathcal{H}_G models is shown in Figs. 3 and 4, respectively.

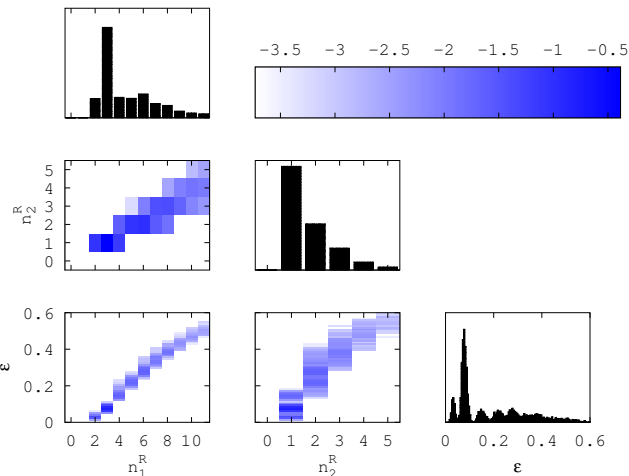


FIG. 3: Marginal posteriors of the singlet charges and ϵ in the model \mathcal{A}_G using the full data set. In the two-dimensional plots, the color scale denotes the base-10 logarithm of the posterior probability of each region.

In Fig. 3 the posteriors of the singlet charges and ϵ in the generalised anarchical model is plotted using priors with $\lambda = 2$ for the mean of the charges and $\epsilon_0 = 10^{-2}$

for the parameter in the prior on ϵ .³ From the three two-dimensional posteriors of the singlet charges and ϵ , one notes the strong degeneracy between the parameters, which follows from Eq. (14) and the fact that the neutrino data is insensitive to the values of ϵ and the charges. Eq. (14) translates into $\epsilon^{n_1^R} \simeq m_e/m_\tau$, $\epsilon^{n_2^R} \simeq m_\mu/m_\tau$, and $n_1^R/n_2^R \simeq \log(m_e/m_\tau)/\log(m_\mu/m_\tau) \simeq 2.9$. The width of these regions are determined by the widths of the $\mathcal{O}(1)$ factors, and the uncertainty of ϵ is as in Sec. III E. One naively expects an exact one-dimensional degeneracy; however, this is broken by the assignment of a prior which gives different posteriors to different points along that curve. These are the prior on the charges and on ϵ , in particular its upper limit which yields upper limits also on the charges. Also the widths of the $\mathcal{O}(1)$ elements affects the viability of any point along the degeneracy differently, since points which require less fine-tuning of the $\mathcal{O}(1)$ elements are inherently favoured. No really strong constraints on any single parameter can be extracted though, although with a high probability $n_1^R \in [2, 8]$, $n_2^R \in [1, 3]$ and $\epsilon < 0.4$.

In the posterior of ϵ there are multiple peaks, each of which is generally dominated by a single pair of charges. The first three peaks correspond to charges (2, 1), (3, 1), and (4, 1), respectively, while the remaining ones which can be seen are dominated for pairs of charges with $n_2^R = 2$.

In Fig. 4 the posteriors of all the charges and ϵ in the generalised hierarchical model is plotted, also using $\lambda = 2$ and $\epsilon_0 = 10^{-2}$. Once again, $n_1^L = n_2^L = 0$ is excluded by the definition of the model.

First, the three two-dimensional posteriors of the singlet charges and ϵ show that the degeneracies have changed shape and position, and gotten wider with respect to the same plots in Fig. 3. These are of course just what is expected from Eq. (14) when the doublet charges are allowed to be non-zero. In particular, the extra suppression from the doublet charges requires ϵ to be larger in \mathcal{H}_G than \mathcal{A}_G for fixed singlet charges, while the ratio n_1^R/n_2^R is smaller, around 2.

Most of the posterior probability is assigned to $(n_1^L, n_2^L) = (1, 0)$, (1, 1), and (2, 1). Two of these combinations were considered in the Sec. III. We note that in general ϵ prefers to be larger in \mathcal{H}_G than in \mathcal{A}_G , and also the appearance of two peaks in its posterior (on top of a large “background”). From the plots of ϵ vs. n_1^L and n_2^L one sees that the peak at around $\epsilon \approx 0.15$ comes mainly from (1, 0), while the peak at around $\epsilon \approx 0.4$ is associated with (1, 1), and (2, 1). This is consistent with the fact that (1, 0) and (2, 1) were the doublet charges of

³ Since we find that the code becomes rather inefficient when the singlet charges are kept as free parameters when analysing this model, we evaluate the posterior by making separate runs with fixed charges and then average the results using the evidence and prior for each charge pair.

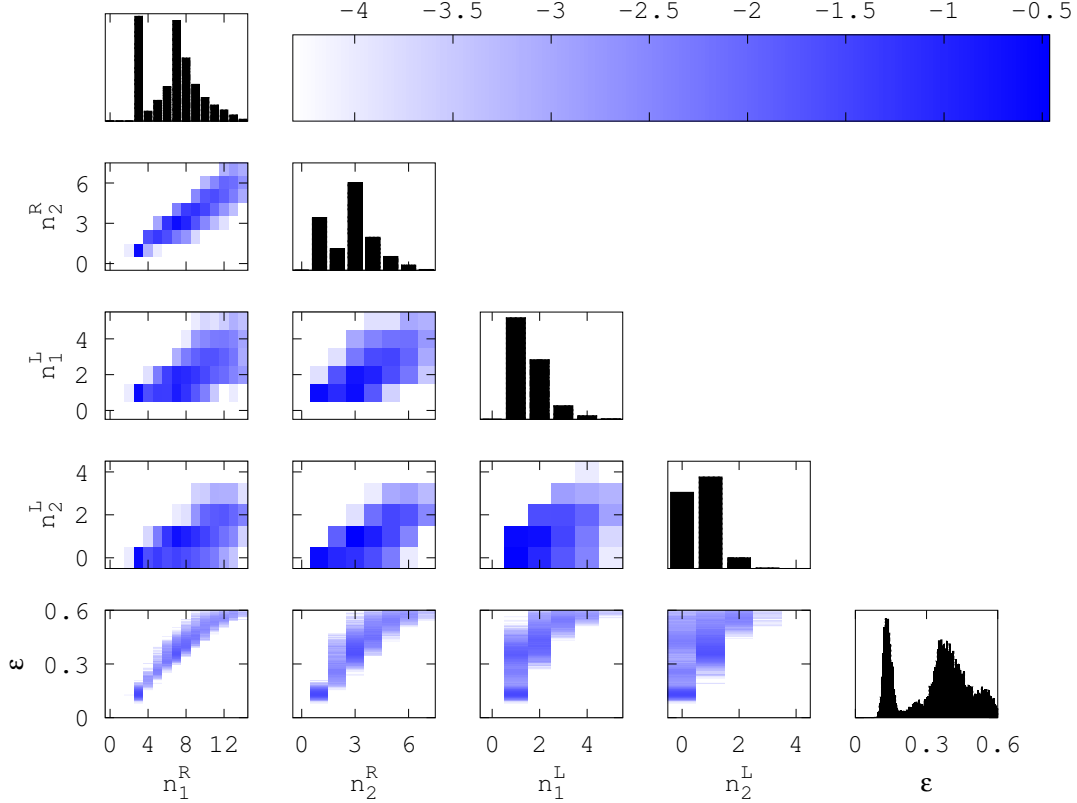


FIG. 4: Marginal posteriors of the singlet charges and ϵ in the model \mathcal{M}_G using the full data set. In the two-dimensional plots, the color scale denotes the base-10 logarithm of the posterior probability of each region.

the $A_{\mu\tau}$ and H/H' models in Sec. III B, and that these different values of ϵ were found to be preferred in Fig. 1. Finally, ϵ has a quite strict lower bound, since for even the smallest possible choice of the doublet charges, a very small ϵ will give a too large hierarchy in the neutrino sector.

Furthermore, the charges for the singlets must typically be larger in \mathcal{H}_G than in \mathcal{A}_G . This mainly follows from the fact that the doublet charges together with ϵ are essentially used to fit four observables (the three mixing angles and the ratio r), while the singlet charges are used to fit only two (the mass hierarchies of the charged leptons). Then, the non-zero doublet charges require ϵ to not be very small so that only a mild hierarchy in the neutrino sector is achieved. Therefore, in order to fit the charged lepton data, larger singlet charges are preferred in \mathcal{H}_G than in \mathcal{A}_G .

C. Results: model comparison

Similarly as in Sec. III B, no parameter constraint can be extracted and no model comparison can be performed from a χ^2 analysis. Indeed for any given value of the charges, and indeed for any fixed tuple $(n_1^R, n_2^R, n_1^L, n_2^L, \epsilon)$,

the data can always be fitted perfectly. The comparison among \mathcal{H}_G and \mathcal{A}_G is then performed on posteriors and evidences.

The evidence of a model M with free charges c is given by

$$\Pr(\mathbf{D}|M) = \Sigma_c \Pr(\mathbf{D}|c, M) \Pr(n|M) = \Sigma_c \mathcal{Z}_c \pi(c) \quad (19)$$

and so the full evidence is the average of the evidences with fixed charges over the prior on the charges. Hence, a model with a large number of unspecified charges can have a small evidence if a large portion of those charges does not predict the data accurately, even if there is some specific combination of charges which predict the data well.

For our default choice of priors with $\lambda = 2$ and $\epsilon_0 = 10^{-2}$, the comparison of \mathcal{H}_G and \mathcal{A}_G yields a Bayes factor

$$\log \mathcal{B} = \log (\mathcal{Z}_{\mathcal{H}_G} / \mathcal{Z}_{\mathcal{A}_G}) = 1.7, \quad (20)$$

which means that the hierarchical model is weakly preferred compared to the anarchical one. The uncertainty on the value reported in the previous equation, as on all the Bayes factors in this section, is about 0.2.

Bayes factors can often depend crucially on the priors employed, in particular when the parameter in question

only appears in one of the models. Since the doublet charges are free parameters only in \mathcal{H}_G , it is then important to check whether the results discussed above are stable under different assumptions on the priors. Usually, the prior on the additional parameters are taken as unrelated to the common parameters, while in the present case the singlet and doublet charges have the same priors, which could result in a smaller prior dependence.

As a first possibility, the priors on the charges have been taken as the geometric distribution (as used in the default case) with the expected value λ of the smallest charges varied between 1 and 3. A second possibility considered is of a uniform prior on the charges up to a maximum of n_{\max} between 7 and 15: this implies that the marginal prior on the largest charge is proportional to its value (up to n_{\max}), and so implicitly larger charges are a priori favoured.

For the prior on ϵ , we have varied the value of ϵ_0 , that in the previous section was fixed at 10^{-2} . Since ϵ is a free parameter of all models, the impact on the evidence is in general expected to be small. As long as ϵ_0 is smaller than the smallest value preferred by any model, it should not effect neither the posteriors nor the Bayes factors, and in fact all posterior inferences should have a unique limit as $\epsilon_0 \rightarrow 0$ (this also applies to the specific models in Sec. III B). $\epsilon_0 = 10^{-2}$ is already small enough for this limit to be well approximated (we have also checked this numerically), and so we do not show any results for smaller ϵ_0 . As $\epsilon_0 \rightarrow \infty$, $\pi(\epsilon)$ becomes a uniform distribution.

We find the following.

Geometric charge priors

For small ϵ_0 the evidence for \mathcal{A}_G does not really depend on λ , which is expected since in this case charges of all magnitudes are roughly equally good (along the degeneracy). However, for ϵ_0 with a uniform prior (i.e. large ϵ_0), larger λ gives larger evidence: indeed large ϵ , and so larger charges, are preferred.

For \mathcal{H}_G , the effect of varying λ is in most cases very small, regardless of the prior on ϵ . If only the prior on the doublet charges is made wider, the evidence decreases (since small charges predict the data better), and vice versa. If only the prior on the singlet charges is made wider, the evidence instead increases, and vice versa. Since the priors are the same on both sets of charges, these effect partially cancel, leaving only small changes in the evidence. Since large ϵ is preferred in \mathcal{H}_G , the uniform prior gives a larger evidence.

When the Bayes factor is calculated, many of the changes in the evidences of the models tend to cancel, giving:

$$\begin{aligned} \log \mathcal{B} &= 1.7 \rightarrow 1.5 & (\lambda : 1 \rightarrow 3, \epsilon_0 < 10^{-2}) \\ \log \mathcal{B} &= 2.3 \rightarrow 2.0 & (\lambda : 1 \rightarrow 3, \epsilon_0 = \infty). \end{aligned} \quad (21)$$

Hence, in total the model comparison is very stable against changes in the priors with weak to “almost moderate” preference of \mathcal{H}_G over \mathcal{A}_G .

Uniform charge priors

For the uniform prior on the charges, much more prior is put on large values of the charges, especially for large n_{\max} . Hence, one expects \mathcal{A}_G to be relatively unaffected by this modification, while the small doublet charges required in \mathcal{H}_G will work against it. Indeed, as expected, we find:

$$\begin{aligned} \log \mathcal{B} &= 1.3 \rightarrow 0.8 & (n_{\max} : 7 \rightarrow 15, \epsilon_0 < 10^{-2}) \\ \log \mathcal{B} &= 1.8 \rightarrow 1.0 & (n_{\max} : \rightarrow 15, \epsilon_0 = \infty). \end{aligned} \quad (22)$$

A uniform prior in the interval $[1/3, 3]$ on the $\mathcal{O}(1)$ parameters has also been considered and it turns out that the difference compared to the standard (Gaussian) case is within the numerical uncertainties.

V. FUTURE PROSPECTS

In this section, we comment on the ability of future low-energy experiments to distinguish between our considered models. If a new data set \mathbf{D}_f is added to the current one, \mathbf{D} , the resulting Bayes factor between two models becomes

$$\mathcal{B}_f = \frac{\Pr(\mathbf{D}_f, \mathbf{D} | M_1)}{\Pr(\mathbf{D}_f, \mathbf{D} | M_0)} = \frac{\Pr(\mathbf{D}_f | \mathbf{D}, M_1) \Pr(\mathbf{D} | M_1)}{\Pr(\mathbf{D}_f | \mathbf{D}, M_0) \Pr(\mathbf{D} | M_0)}, \quad (23)$$

i.e., the new Bayes factor equals the present one multiplied by $\mathcal{B}_{\text{upd}} = \Pr(\mathbf{D}_f | \mathbf{D}, M_1) / \Pr(\mathbf{D}_f | \mathbf{D}, M_0)$, with evidences given by

$$\Pr(\mathbf{D}_f | \mathbf{D}, M) = \int \Pr(\mathbf{D}_f | \boldsymbol{\Theta}, M) \Pr(\boldsymbol{\Theta} | \mathbf{D}, M) d\boldsymbol{\Theta}. \quad (24)$$

Hence, the change in the Bayes factor by the addition of a new set of data is given by calculating the average of the future likelihood over the present-day posterior. As an ideal situation, consider the case where the new data determines some combinations of parameters exactly (for example, some set of low-energy observables), so that $\Pr(\mathbf{D}_f | \boldsymbol{\Theta}, M) = \delta(\alpha(\boldsymbol{\Theta}) - \alpha_0)$. Then

$$\frac{\Pr(\mathbf{D}_f | \mathbf{D}, M_1)}{\Pr(\mathbf{D}_f | \mathbf{D}, M_0)} = \frac{\Pr(\alpha_0 | \mathbf{D}, M_1)}{\Pr(\alpha_0 | \mathbf{D}, M_0)}, \quad (25)$$

and therefore if a perfect measurement of a single observable is to be able to increase the evidence of, say, 5 log units, then the ratio of the current posteriors at the true value must differ by a factor of $e^5 \simeq 150$.

Regarding the data adopted here, all the functions of parameters used to constrain the models are well-constrained, so that the posteriors of those parameters follow the experimental likelihoods closely in all the considered models. Hence, improved measurements of those

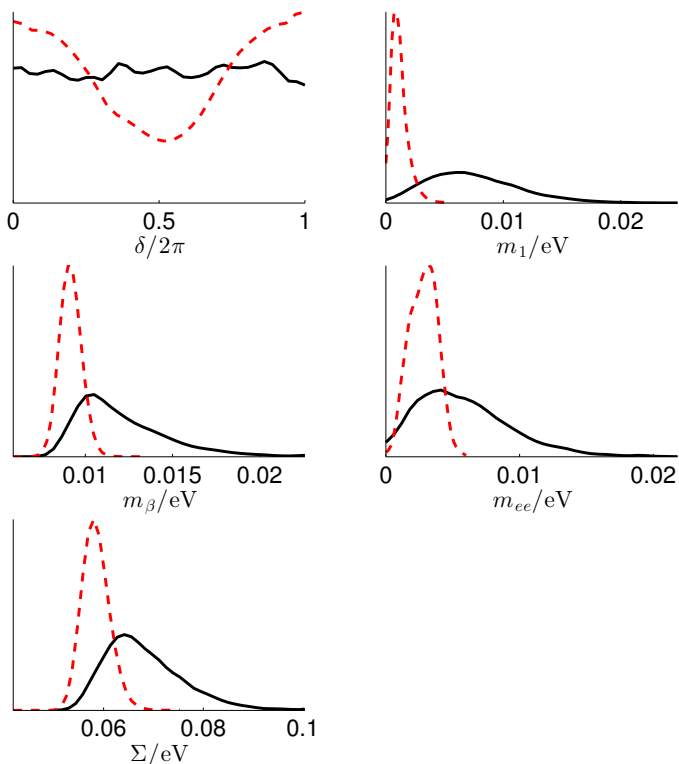


FIG. 5: Posteriors for the Dirac CP-phase δ , the lightest neutrino mass m_1 , the effective $0\nu 2\beta$ -decay effective mass m_{ee} , the β -decay mass m_β , and the sum of the neutrino masses Σ , for A' (black-continuous line) and H' (red-dashed line).

parameters can not further discriminate between the models.

There are other observables which could be accurately measured in future experiments, and in principle could be used to distinguish between the models. These are primarily the CP-phase δ and observables related to the values of neutrino masses (m_{ee} , m_β , Σ). We plot the posteriors of these variables as well as the lightest neutrino mass m_1 in Fig. 5 for the model A' and H' . Similar posteriors are expected for the other hierarchical and anarchical models. Correlations among (m_{ee} , m_β , Σ) and m_1 can be read in the two-dimensional posterior plots, displayed in Fig. 6.

A precise measurement of δ can only give a very minor further discrimination between the models. On the other hand, a precise measurement of the sum of neutrino masses Σ at about 0.1 eV could give in principle strongly increased support for anarchical models; similar arguments hold for the other variables. Contrary, very stringent upper limits on the different observables could give support to the hierarchical models. However, the practical feasibility of these measurements are not very good in the near future.

As discussed earlier, only the normal mass ordering for the neutrino spectrum has been considered here since it is rather strongly preferred in both the anarchical and

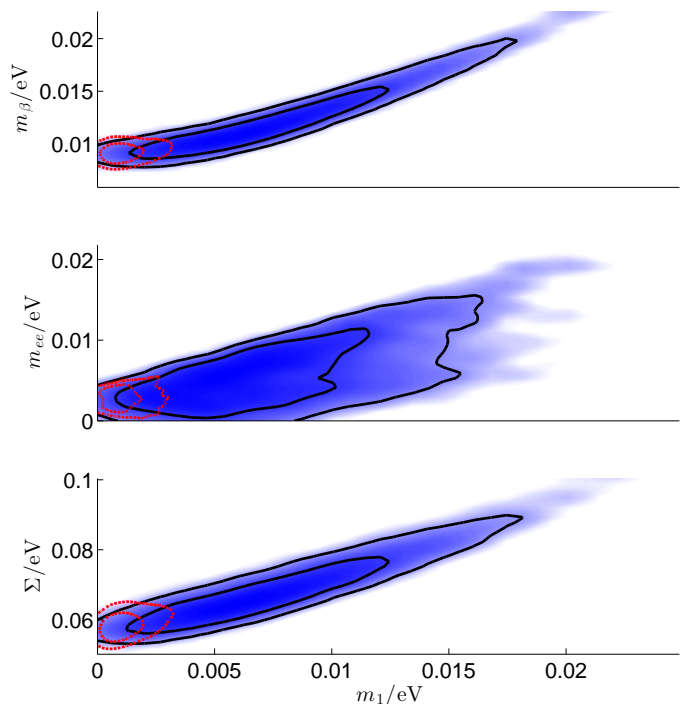


FIG. 6: The two-dimensional posterior of m_1 versus m_{ee} , m_β , and Σ . The contours represent the 1σ and 2σ confidence level regions for A' (black) and H' (red).

hierarchical models. However, if definite evidence for the inverted mass ordering would emerge, all the models considered here would be disfavoured in comparison with their inverted-ordering counter parts. After that, the comparison would be between inverted hierarchical and inverted anarchical models.

VI. CONCLUSIONS

In this letter, a conclusive comparison among anarchical and more symmetric approaches to explain the flavour puzzle in the leptonic sector is presented. The method used is based on Bayesian inference and has been applied to a series of flavour models whose effective Lagrangian shows an invariance under the Abelian group $U(1)$. Two distinct setups have been considered: i) in the first one, the $U(1)$ -charges of the SM fields and of the additional scalar field responsible for the flavour symmetry breaking mechanism are fixed to specific values, determined through phenomenological considerations or by a statistical analysis; ii) in the second one, the $U(1)$ -charges, or part of them, are kept as free parameters and the statistical procedure determined those ones, for which the model best reproduces the data, limiting fine-tunings. Case i) corresponds of giving to the anarchical models A and A' the same prior probability with respect to constructions with built-in hierarchies among the matrix elements, $A_{\mu\tau}$, H and H' . On the other hand, in case ii)

the comparison is among the class of models embedding the anarchical ansatz \mathcal{A}_G and that of models based on a more symmetric principle \mathcal{H}_G . In both cases, models with hierarchical matrix elements are preferred over the anarchical ones and the only difference among the two approaches is the precise values of the evidence: almost moderate in case i), and weak in case ii). This study confirms and extends the previous results in Ref. [44].

The stability of the results have been checked modifying the most sensible aspect of Bayesian inference, i.e. the priors on the parameters of the models. The prior of ϵ has been essentially taken to be either log-uniform or uniform. The priors of the $\mathcal{O}(1)$ parameters entering the Yukawa matrices follow either a distribution constructed from the Haar measure or a uniform distribution in the interval $[1/3, 3]$. Finally, the priors of the charges, when taken as free parameters, have chosen with either a geometric or a uniform distribution in a given support. The results show a slight dependence on the choice of priors, that however do not significantly change the conclusions.

Improvements on the precision of the used data or the addition of new data related to the neutrino mass spectrum and mixing matrix, such as m_{ee} , m_β , Σ and δ , will most likely a very minor effects on the results presented here. On the other hand, understanding the neutrino spectrum ordering could have an impact on the analysis: here the focus was only on the normal ordering, as the inverted one is typically linked to a maximal value of the solar angle, in contrast with the observations.

This work shows the power of the Bayesian inference in comparing models and extracting information on the

model parameters. The only constructions considered here are based on the Abelian $U(1)$ flavour symmetry and with neutrino masses described by the effective Weinberg operator. Extending the analysis to the case where a See-Saw mechanism explains the lightness of the active neutrino masses is straightforward. On the other hand, to include in the comparison models based on different symmetries than $U(1)$, such as non-Abelian continuous or discrete ones, is a non trivial task and is left for future studies.

Acknowledgements

We thank Mattias Blennow, Enrique Fernandez Martinez and Tommy Ohlsson for participating at the initial phase of this project. J.B. and L.M. acknowledge partial support by European Union FP7 ITN INVISIBLES (Marie Curie Actions, PITN-GA-2011-289442). D.M. acknowledges MIUR (Italy) for financial support under the program Futuro in Ricerca 2010 (RBFR10O36O). L.M. acknowledges partial support by the Juan de la Cierva programme (JCI-2011-09244) and by the Spanish MINECOs “Centro de Excelencia Severo Ochoa” Programme under grant SEV-2012-0249. L.M. thanks the Göran Gustafsson Foundation for financial support and KTH Royal Institute of Technology, the Dipartimento di Matematica e Fisica dell’Università Roma Tre and the theoretical group at CERN for hospitality during the development of this project.

-
- [1] **T2K Collaboration**, K. Abe *et. al.*, Phys. Rev. Lett. **107** (2011) 041801, [[arXiv:1106.2822](#)].
 - [2] **T2K Collaboration** Collaboration, K. Abe *et. al.*, Phys.Rev.Lett. **111** (2013) 211803, [[arXiv:1308.0465](#)].
 - [3] **T2K Collaboration** Collaboration, K. Abe *et. al.*, [[arXiv:1311.4750](#)].
 - [4] **MINOS Collaboration**, P. Adamson *et. al.*, Phys. Rev. Lett. **107** (2011) 181802, [[arXiv:1108.0015](#)].
 - [5] **MINOS Collaboration** Collaboration, P. Adamson *et. al.*, Phys.Rev.Lett. **110** (2013), no. 17 171801, [[arXiv:1301.4581](#)].
 - [6] **MINOS Collaboration** Collaboration, P. Adamson *et. al.*, Phys.Rev.Lett. **110** (2013) 251801, [[arXiv:1304.6335](#)].
 - [7] **DOUBLE-CHOOZ** Collaboration, Y. Abe *et. al.*, [[arXiv:1112.6353](#)].
 - [8] **Double Chooz Collaboration** Collaboration, Y. Abe *et. al.*, Phys.Rev. **D86** (2012) 052008, [[arXiv:1207.6632](#)].
 - [9] **DAYA-BAY** Collaboration, F. P. An *et. al.*, [[arXiv:1203.1669](#)].
 - [10] **Daya Bay Collaboration** Collaboration, F. An *et. al.*, [[arXiv:1310.6732](#)].
 - [11] **RENO** Collaboration, J. K. Ahn *et. al.*, [[arXiv:1204.0626](#)].
 - [12] G. L. Fogli, E. Lisi, A. Marrone, A. Palazzo, and A. M. Rotunno, Phys. Rev. **D84** (2011) 053007, [[arXiv:1106.6028](#)].
 - [13] T. Schwetz, M. Tortola, and J. W. F. Valle, New J. Phys. **13** (2011) 109401, [[arXiv:1108.1376](#)].
 - [14] P. Machado, H. Minakata, H. Nunokawa, and R. Zukanovich Funchal, JHEP **1205** (2012) 023, [[arXiv:1111.3330](#)].
 - [15] M. Tortola, J. Valle, and D. Vanegas, [[arXiv:1205.4018](#)].
 - [16] G. Fogli, E. Lisi, A. Marrone, D. Montanino, A. Palazzo, *et. al.*, [[arXiv:1205.5254](#)].
 - [17] M. Gonzalez-Garcia, M. Maltoni, J. Salvado, and T. Schwetz, JHEP **1212** (2012) 123, [[arXiv:1209.3023](#)]. Updates at <http://www.nu-fit.org>.
 - [18] F. Capozzi, G. Fogli, E. Lisi, A. Marrone, D. Montanino, *et. al.*, [[arXiv:1312.2878](#)].
 - [19] G. Altarelli and F. Feruglio, New J. Phys. **6** (2004) 106, [[hep-ph/0405048](#)].
 - [20] R. N. Mohapatra and A. Y. Smirnov, Ann. Rev. Nucl. Part. Sci. **56** (2006) 569–628, [[hep-ph/0603118](#)].
 - [21] W. Grimus, PoS **P2GC** (2006) 001, [[hep-ph/0612311](#)].
 - [22] M. C. Gonzalez-Garcia and M. Maltoni, Phys. Rept. **460** (2008) 1–129, [[arXiv:0704.1800](#)].
 - [23] G. Altarelli, Nuovo Cim. **C32N5-6** (2009) 91–102, [[arXiv:0905.3265](#)].
 - [24] G. Altarelli and F. Feruglio, Rev. Mod. Phys. **82** (2010) 2701–2729, [[arXiv:1002.0211](#)].

- [25] W. Grimus and P. O. Ludl, [arXiv:1110.6376](#).
- [26] G. Altarelli, F. Feruglio, and L. Merlo, *Fortsch.Phys.* **61** (2013) 507–534, [[arXiv:1205.5133](#)].
- [27] F. Bazzocchi and L. Merlo, *Fortsch.Phys.* **61** (2013) 571–596, [[arXiv:1205.5135](#)].
- [28] S. Morisi and J. Valle, *Fortsch.Phys.* **61** (2013) 466–492, [[arXiv:1206.6678](#)].
- [29] S. F. King and C. Luhn, *Rept.Prog.Phys.* **76** (2013) 056201, [[arXiv:1301.1340](#)].
- [30] G. Altarelli, [arXiv:1304.5047](#).
- [31] S. F. King, A. Merle, S. Morisi, Y. Shimizu, and M. Tanimoto, [arXiv:1402.4271](#).
- [32] S. King and G. G. Ross, *Phys.Lett.* **B520** (2001) 243–253, [[hep-ph/0108112](#)].
- [33] S. King and G. G. Ross, *Phys.Lett.* **B574** (2003) 239–252, [[hep-ph/0307190](#)].
- [34] I. de Medeiros Varzielas, S. F. King, and G. G. Ross, *Phys. Lett.* **B648** (2007) 201–206, [[hep-ph/0607045](#)].
- [35] S. Davidson and F. Palorini, *Phys. Lett.* **B642** (2006) 72–80, [[hep-ph/0607329](#)].
- [36] B. Grinstein, V. Cirigliano, G. Isidori, and M. B. Wise, *Nucl. Phys.* **B763** (2007) 35–48, [[hep-ph/0608123](#)].
- [37] R. Alonso, G. Isidori, L. Merlo, L. A. Munoz, and E. Nardi, *JHEP* **06** (2011) 037, [[arXiv:1103.5461](#)].
- [38] R. Alonso, M. Gavela, D. Hernandez, and L. Merlo, *Phys.Lett.* **B715** (2012) 194–198, [[arXiv:1206.3167](#)].
- [39] R. Alonso, M. Gavela, D. Hernandez, L. Merlo, and S. Rigolin, *JHEP* **1308** (2013) 069, [[arXiv:1306.5922](#)].
- [40] R. Alonso, M. Gavela, G. Isidori, and L. Maiani, [arXiv:1306.5927](#).
- [41] G. Altarelli, F. Feruglio, and I. Masina, *JHEP* **11** (2000) 040, [[hep-ph/0007254](#)].
- [42] G. Altarelli, F. Feruglio, and I. Masina, *JHEP* **01** (2003) 035, [[hep-ph/0210342](#)].
- [43] W. Buchmuller, V. Domcke, and K. Schmitz, *JHEP* **03** (2012) 008, [[arXiv:1111.3872](#)].
- [44] G. Altarelli, F. Feruglio, I. Masina, and L. Merlo, *JHEP* **1211** (2012) 139, [[arXiv:1207.0587](#)].
- [45] G.-J. Ding, S. Morisi, and J. Valle, *Phys.Rev.* **D87** (2013), no. 5 053013, [[arXiv:1211.6506](#)].
- [46] L. J. Hall, H. Murayama, and N. Weiner, *Phys. Rev. Lett.* **84** (2000) 2572–2575, [[hep-ph/9911341](#)].
- [47] N. Haba and H. Murayama, *Phys. Rev.* **D63** (2001) 053010, [[hep-ph/0009174](#)].
- [48] A. de Gouvea and H. Murayama, *Phys. Lett.* **B573** (2003) 94–100, [[hep-ph/0301050](#)].
- [49] A. de Gouvea and H. Murayama, [arXiv:1204.1249](#).
- [50] J. Espinosa, [hep-ph/0306019](#).
- [51] E. T. Jaynes, *Probability theory: The logic of science*. Cambridge University Press, 2003.
- [52] D. S. Sivia and J. Skilling, *Data analysis: a Bayesian tutorial*. Oxford University Press, 2006.
- [53] T. J. Lored, Maximum-Entropy and Bayesian Methods (1990) 81–142.
- [54] R. Trotta, *Contemp.Phys.* **49** (2008) 71–104, [[arXiv:0803.4089](#)].
- [55] M. Hobson *et. al.*, eds., *Bayesian methods in cosmology*. Cambridge University Press, 2010.
- [56] F. Feroz, B. C. Allanach, M. Hobson, S. S. AbdusSalam, R. Trotta, *et. al.*, *JHEP* **0810** (2008) 064, [[arXiv:0807.4512](#)].
- [57] S. AbdusSalam, B. Allanach, M. Dolan, F. Feroz, and M. Hobson, *Phys. Rev.* **D80** (2009) 035017, [[arXiv:0906.0957](#)].
- [58] J. Bergstrom, *JHEP* **1208** (2012) 163, [[arXiv:1205.4404](#)].
- [59] J. Bergstrom, *JHEP* **1302** (2013) 093, [[arXiv:1212.4484](#)].
- [60] H. Jeffreys, *Theory of probability*. Oxford University Press, 1961.
- [61] R. E. Kass and A. E. Raftery, *J. Am. Stat. Ass.* **90** (1995) 1773.
- [62] F. Feroz and M. Hobson, *Mon. Not. Roy. Astron. Soc.* **384** (2008) 449, [[arXiv:0704.3704](#)].
- [63] F. Feroz, M. Hobson, and M. Bridges, *Mon. Not. Roy. Astron. Soc.* **398** (2009) 1601–1614, [[arXiv:0809.3437](#)].
- [64] F. Feroz, M. Hobson, E. Cameron, and A. Pettitt, [arXiv:1306.2144](#).
- [65] C. D. Froggatt and H. B. Nielsen, *Nucl. Phys.* **B147** (1979) 277.
- [66] R. E. Kass and L. Wasserman, *J. Am. Stat. Ass.* **91** (1996) 1343.
- [67] Y. Bai and G. Torroba, *JHEP* **1212** (2012) 026, [[arXiv:1210.2394](#)].